HM Government

Simplifying the transition to Individual Electoral Registration

An evaluation of the 'confirmation dry run' – using data matching to confirm electors on the electoral register

Henry Grigg, Will Lockhart and Maria Hannan. October 2013

Cabinet Office
25 Great Smith Street
London SW1P 3BQ

Publication date: October 2013

Any enquiries regarding this document/publication should be sent to us at
electoralregistration@cabinet-office.gsi.gov.uk

Acknowledgements

The authors would like to thank all 380 local authorities who took part in the Confirmation Dry Run for their hard work and enthusiasm, as well as the information they provided to aid the evaluation. Similarly we would like to thank the Electoral Management System providers for supporting the local authorities and aiding the evaluations. Thanks are also due to the Electoral Commission for their collaborative approach to the evaluation, in particular Phil Thompson and Davide Tiberti.

Contents

Ministerial foreword

The electoral register is a key building block of our democracy. The Government sees both registering to vote and voting as civic duties and we strongly encourage people to do both. The introduction of Individual Electoral Registration (IER) in Great Britain will modernise the way people register to vote, help to tackle electoral fraud and improve confidence in the electoral register. Registration will become an individual responsibility rather than an act carried out by just one person in each household.

The transition to IER is scheduled to begin in 2014 in England, Wales and Scotland in line with the Implementation Plan we published in summer 2012, although this timing will be confirmed later this year. In Scotland, the transition will be delayed until after the Independence Referendum on 18 September so that these two important events do not overlap and potentially cause confusion for electors. Any new applications made after the transition has begun will be an IER application and must include the applicant's date of birth (DOB) and National Insurance Number (NINO) so that their identity can be verified, increasing the integrity of the register.

The first major step in transition will be the matching of existing electoral registers against records held by the Department for Work and Pensions – 'confirmation'. This report sets out the findings from a complete national test carried out in the summer of 2013 of the process which will automatically confirm the majority of electors on the register during the transition to IER in 2014. The test (the confirmation dry run) involved the matching of all 380 registers, with around 46m people, against Department for Work and Pensions data, using the process, IT, and people who will do it for real next year. The results are very positive: 78% of electors matched, higher than achieved by our previous pilots in 2012 and 2011, and local data matching has the potential to add an average of 7%. This means that most electors will not need to apply under IER unless their circumstances change (e.g. they move house), and that the risk of a drop in the register during transition is therefore significantly reduced. However, we are aware that there is still work to do to register the remaining electors who will not be confirmed as well as those who are currently not registered.

We have identified lessons to be learned from this exercise, but it is a positive example of collaboration between the Cabinet Office, Government Digital Service, Department for Work and Pensions and Electoral Registration Officers (and administrators), their suppliers, using data to simplify the citizens' experience of public services. It has also shown the success of the IT system in place to help ensure the change happens as smoothly as possible.

Chapter 1
Introduction

The current system of electoral registration is based on an annual household canvass sent to each address, which is completed by one individual on behalf of everyone who lives in the house. In 2014 this will be replaced by a system of Individual Electoral Registration (IER) and individuals will have to provide personal identifiers such as National Insurance Numbers (NINOs) and dates of birth (DOBs) to register.

One of the key aims of the Government is to ensure that the electoral register remains as complete and accurate as possible under IER. The Cabinet Office have conducted a series of data matching pilots since 2011 and these have identified the use of data held by the Department for Work and Pensions (DWP) to confirm individuals currently on the electoral register without requiring them to provide personal identifiers – these people can be 'passported' across to the new system. This will allow electoral administrators to focus their limited resources on the minority of electors who cannot be confirmed as well as those currently not registered.

Pilots conducted in 2011 suggested that 66% of existing electors might be confirmed using this process. However, those pilots did not set out to test confirmation and so further pilots were undertaken in 2012 to specifically test data matching for the purposes of confirmation and to check the accuracy of the data. These pilots found that around 70% of electors could be confirmed. The pilots also found that the vast majority of electors who were matched in the pre-canvass register (95%) were subsequently confirmed as resident at the same address during the annual canvass – showing that we can be confident in the accuracy of the data.

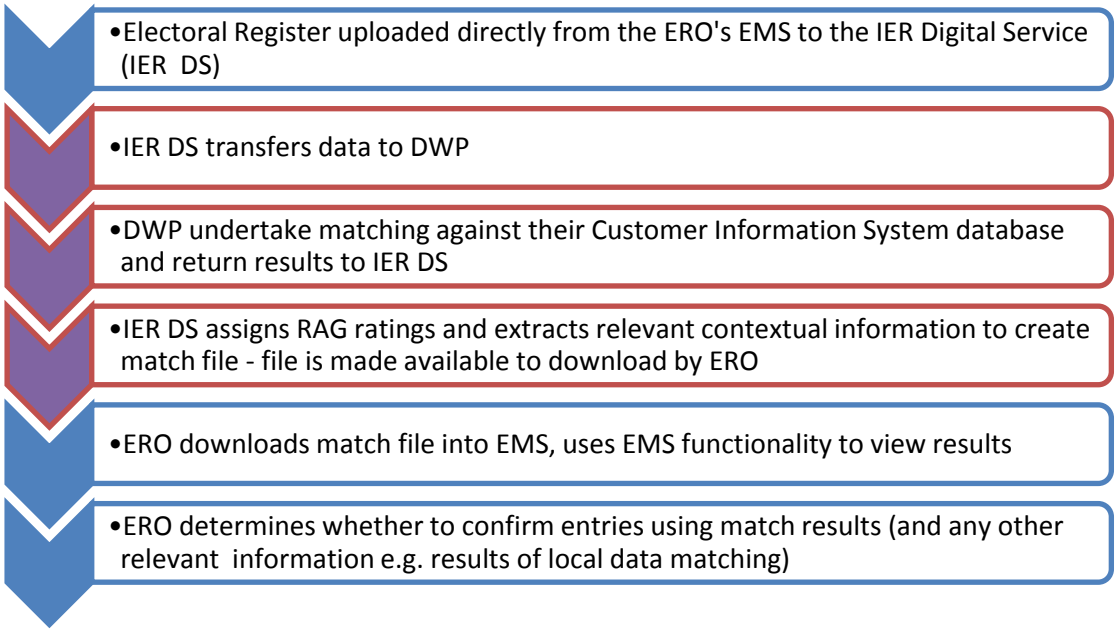
The confirmation pilots in 2012 took place in 14 areas and were a chance to develop the matching algorithm – working with both technical experts at DWP and five 'beacon' local authorities – and test the accuracy of the data. They were not, however, able to fully test the IER Digital Service (IER DS) as the connections to transfer the data from the local authorities to the IER DS and from the IER DS to DWP were not ready (and were not planned to be at that stage). The data was therefore transferred, as planned, to and from the pilot areas by secure courier and was sent as CSV files rather than via reports within Electoral Management Software (EMS) – meaning that Electoral Registration Officers (EROs) were required to

analyse the data independently as opposed to using reporting functionality in their software. The Confirmation Dry Run (CDR) was therefore planned and conducted during the summer of 2013. This was an opportunity to test a fully IT enabled dry run of the confirmation process ahead of it happening in a live environment in 2014.

Chapter 2
Methodology

During July and October 2013 every Local Authority (LA) in England and Wales and all Scottish Valuation Joint Boards (VJBs) were required to participate in CDR. The figure below sets out the process for matching.

Figure 1: Outline confirmation process (Steps in light blue require no action from the ERO)



Notes: Steps in light blue require no action from the ERO. In live-running, once the steps outlined above are complete local areas will carry-out follow-up action, including writing to individuals who have been confirmed to notify them their details have been transferred to the IER register and issuing invitations to register for those individuals who have not been confirmed.

All areas were scheduled to upload their registers on a specific day and could then download their match results 5 working days afterwards. The results were available in their EMS system which then allowed electoral administrators to view their match results – electors were given a red, amber or green rating.[1] A green match indicated a positive result, amber indicated a possible match and a red match indicated that no

[1] Details on the matching process and algorithm can be found at Annex A.

match could be found. Some contextual information was also provided to inform administrators what for example had failed, e.g. address.

Administrators had the option of conducting additional local data matching if they chose. This had the potential to confirm additional electors using local sources of data such as council tax or housing benefit data. They could also check DWP green matches if they chose to but this was used less frequently.

Reports were produced within the EMS system and sent to the Cabinet Office (and subsequently the Electoral Commission (EC) for their independent evaluation) to provide statistics on match rates, broken down by elector type and wards. Additional reports were sent where an administrator had conducted local data matching. The Cabinet Office and EC also designed a joint survey which administrators were requested to complete to provide feedback on the process of CDR, their views on the matching and how they might use the data in a live environment, as well as information on local data matching and how things could be improved in the future. The Cabinet Office, EC and EMS providers also provided guidance to administrators on how to interpret their results and conduct their CDR process. The results of CDR will help local authorities and Valuation Joint Boards to plan their workloads for next year and will help the Cabinet Office award funding to different areas.

It should be noted that all percentages presented in this report have been rounded.

Chapter 3
Process of CDR

The survey on which the largely qualitative results presented in this section are drawn was completed by 320 LAs and VJBs, a response rate of 80%. Data on the match rate of all authorities is presented below in chapter 4.

Preparatory guidance

To help prepare administrators for CDR a variety of guidance was produced. 93% of respondents felt that the guidance they received from the Cabinet Office was very or fairly helpful in preparing them for the CDR. 93% of respondents found the guidance issued by the EMS/in-house provider to be very of fairly helpful, and 78% of respondents found very or fairly helpful the guidance issued by the EC. However, 52 respondents (16%) felt that the volume of guidance they received in a short period from three different organisations – Cabinet Office, the EC and their EMS supplier – meant that it was impossible to follow. There is a clear call for single, comprehensive guidance documents to be published well in advance of the process and online, rather than 'piecemeal' and by different organisations. Other respondents felt that there could have been clearer guidance issued on how records get their RAG ratings, and step-by-step instructions as to how to go about data matching. Both of these topics are covered in detail by Ministerial guidance, which was published in October 2013.

Data upload

93% of those LAs and VJBs that completed the CDR survey found it very or fairly easy to send the register for matching. The most commonly encountered problems with sending the register were the slow speed of the upload and local IT issues, including problems with local firewalls interfering with the upload or network failures interrupting it. A handful of LAs/VJBs reported that they had encountered problems with their bearer tokens, the authentication credential which acts as the 'key' to secure data transmission between the local authority and the IER DS. However, they were quickly resolved by revoking and activating the tokens. A number of users felt that the instructions issued to them were not clear (dealt with further below).

Some respondents were affected by a problem with duplication of data. From time to time, the IER DS received duplicate batches of up to 500 records from LAs. The most likely cause for this was that the EMS resent the records after a timeout. Prior to CDR, in discussion with the EMS providers, the Government Digital Service (GDS) removed the duplicate constraint, therefore allowing these records. To help alleviate the problem, EMS providers released a software update which allowed the LAs to retrieve all their records.

Data download

82% of respondents were very or fairly satisfied with the way that matched data was returned into the EMS system. The most commonly (and quite widely) reported problem was the slow download speed, with two authorities reporting download times of over 20 hours. This was caused by a combination of problems with aggregated gateways within the GCSX network and local network/IT infrastructure at the LA end. Some respondents felt that they were not given enough notice from the Cabinet Office that their data was ready for download, and would have appreciated a definite time slot having been agreed in advance.

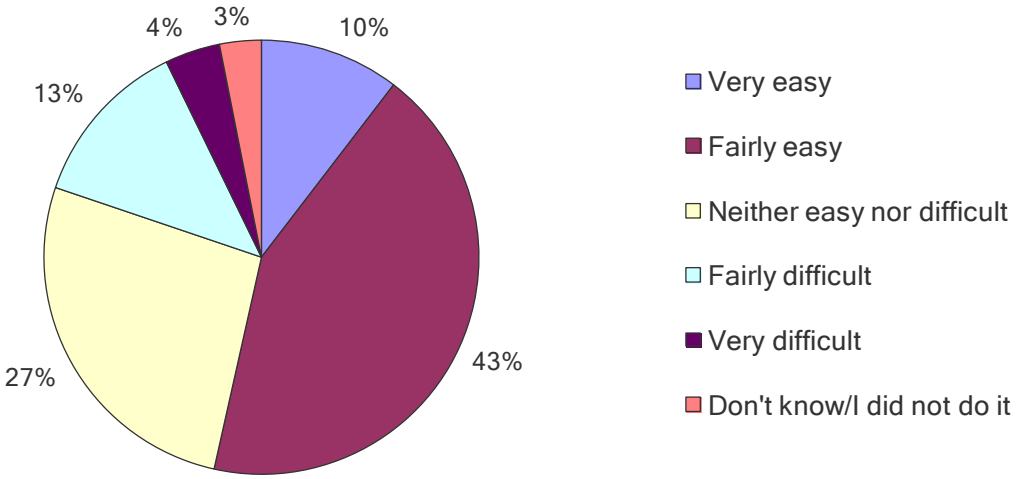
Authorities who uploaded their data before the 15 July encountered a presentational issue in the way that address matches were reported. In those cases where a name was returned as a red, failed match, it was also reported as not having matched on address. This was particularly problematic with regards those households with multiple electors, one of whom returned as green, because the address would then show as a green match. This meant that Electoral Service Managers (ESMs) were confronted with conflicting information about whether the address format they had was the same as one on record in DWP. The issue was addressed during the CDR by altering the matching algorithm, with the result that those electors who matched on address but not name would return a green and a red match respectively. Although this did not change the authority's overall match rate, all 87 affected authorities were offered the chance to upload again, and 21 authorities made use of the offer.

93% of respondents are very or fairly confident in the security of the system for transferring data.

Data analysis

The electoral services team were primarily responsible for analysing the CDR data in 94% of responding authorities. 1% of respondents could not identify the match rate for the individual wards, with 90% able to, and 9% not having tried. Figure 22 shows that 54% of respondents found it very or fairly easy to analyse the data.

Figure 2: How easy or difficult did you find it to analyse the data?



Respondents cited a number of problems they found in analysing the data. The varied skill set of ESMs is likely to be one of the reasons that some respondents found it harder than others to analyse the data returned to them. The most frequently cited reason, however, was a lack of polling district data, and many respondents believe that without this data, the information on their matches is significantly less useful than it otherwise could be for planning purposes. Many respondents would like the reason for the mismatch provided alongside red and amber matches, and at least whether a record failed on name or address; respondents similarly felt that this would help them in clarifying what they felt to be obviously incorrect red matches, such as electors marked as deceased who according to local data are still alive. Some respondents asked for clearer guidance on how to analyse their data, and some indicated that they had not understood the weak flag, amber matches, or the process of data matching as a whole.[2] A number of these respondents expressed frustration that they could not access 'plain English' explanations of the CDR. Finally, a number of respondents were disappointed that data matching was being emphasised as an important part of both CDR and confirmation when it was considered by them to be resource-heavy and not essential to delivering IER.

Respondents suggested a number of ways in which the data analysis could be made easier. The most frequently requested piece of data that respondents would like to see included with the match results is a statement of a reason that a record received a red or amber match, including whether it was the address or name that failed. Other commonly requested pieces of information are: the reason for a fail on the address (i.e., which line of the address failed and why); a breakdown by polling district; provision of the actual information held in the Customer Information System

[2] Where an individual cannot be matched at their precise address on the electoral register, at its most basic level the matching process looks to see if the named elector can be found within their postcode. If this is possible, the 'best match' field will show the word 'Weak' and the overall match result will show as Amber – even where their forename, middle name and surname are exact matches and the address they have provided shows as having been matched by unique property reference number (UPRN). This is because whilst an individual with that name has been found in the near area we cannot be completely confident that they reside at that exact property and/or that it is the same person. (Testing of these records in the pilots showed that whilst the majority of matches were accurate the accuracy level was lower than other matches. We opted to still highlight these records as some areas may wish to use this information to help prioritise which records they match using local data, should they choose to use it.)

(CIS) database at DWP; and a separate list of mixed RAG households. Mixed households arise when some electors in a household are returned as green, while others are returned as red or amber. Respondents – used to a household system of registration – were concerned by these, although it is not uncommon for different residents of the same address to have different RAG statuses based on their different records at DWP. Cabinet Office must clearly communicate to ESMs the reason that these mixed households emerge and that they are to be expected in an individual electoral system. It will also be important that communication to electors during the confirmation write out in 2014 is clear in explaining why individuals in the same property may have different results in terms of their confirmation to avoid any confusion. The Cabinet Office and EC are working together to ensure that there are clear messages to both electoral administrators and electors. Respondents have indicated that a separate list of these households would be helpful for resource planning. Respondents were very clear that they need step-by-step instructions on how to analyse results and carry out local data matching. A number of respondents asked if percentages could be included in the report to save them having to add them themselves.

The Cabinet Office are not able to provide ESMs with the specific field on which the record failed to match at DWP. This approach was tested in pilot studies, but the EROs who took part in the study felt that the resulting match report was too detailed, because the matching algorithm checks each record in 130 different ways and would have to report a match or a fail on each of them. The reporting system as it stands provides a simple mechanism for assessing the relative strength of the matches whilst still providing more detailed information via the best match field and related guidance.

The Cabinet Office are not able to reproduce in the match report the data within CIS against which DWP are matching the electoral record because sharing this information is not proportionate under data protection laws for the purposes of electoral registration and is not permitted by legislation.

Cabinet Office will work closely with the EC on the messaging to electors about the confirmation process to try and ensure that they are aware that no data matching process is perfect and that therefore there will always be some individuals who perhaps should have been matched who were not. Furthermore the Cabinet Office will continue to work with electoral administrators around their understanding of the matching process and minimise the referral of electors to DWP or Her Majesty's Revenue & Customs (HMRC).

Cabinet Office will look into the possibility of providing a breakdown of match results by polling district, and are already working with EMS providers, DWP and GDS to produce a separate list of mixed RAG households, to provide percentages in the downloaded reports and to improve their presentation more generally – particularly on 'weak' address matches. Guidance in plain English on RAG statuses and step-by-step recommendations as to how to conduct local data matching are covered by Ministerial guidance on IER, which was made available to administrators in October 2013.

Support and resourcing

67% of respondents received support from local authority staff outside the immediate electoral administration team to carry out the CDR. 76 respondents received help from their IT department, 23 from staff in their council tax department, 11 from their geographic information department, 7 from their local land and property gazetteer department, and others from the audit, revenues and benefits, fraud, and research teams. IT departments helped with the broadest range of tasks, including applying software patches, updating servers, monitoring connectivity, turning off proxy settings on firewalls, repairing secure connections, and data matching. Colleagues in other departments helped with data matching by preparing and exporting databases for comparison, or by matching itself.

According to nearly all respondents, one of the most time-consuming tasks was downloading the match results, and although officials did not need to be in attendance during the process, they were required to monitor the download, which also left the computer in question busy. Processing and chasing non-matches during local data matching, whether automatically or manually, was also reported as time consuming. A much smaller number of respondents felt that learning about the processes involved with CDR had taken the greatest amount of time, and referenced the large amount of material they had been sent by Cabinet Office.

52% of respondents feel that they do not currently have the resources necessary to conduct the work with the data required for the live confirmation process. Of these respondents, nearly all refer to either the time that they have given to preparing for and understanding the CDR, the time it takes to carry out data matching, or a mixture of the two. Many refer to engaging additional staff or purchasing software upgrades to automate the process, though some express concern that they will need extra funding in order to support these. The Cabinet Office is committed to funding any additional burden the transition to IER places on electoral administrators and will release funding in October 2013.

With regards all concerns about the extra work created by confirmation and data matching in general, it should be borne in mind that the process automatically transfers the majority of the current electorate onto the new IER register reducing both the time and money spent on the write-out and canvassing.

52% of respondents needed to contact the CDR control centre during the process. Of those respondents, 92% found their contact with the CDR control centre in the Cabinet Office to be very or fairly helpful. Respondents felt that the service could be improved through the allocation of a dedicated contact individual in the control centre for each authority, and being called back when they were told that they would be. The Cabinet Office will look into the advantages and disadvantages of providing a dedicated contact individual and thus improving call-back performance.

72 respondents took the opportunity to reflect more broadly on the CDR as a whole. The most common theme (19 responses) was that the CDR had gone better than expected, and many respondents commented on their high match rate and the quality of the support from the Cabinet Office or their Regional Delivery Manager. As one respondent explained,

‘*I think the CDR exercise, for us at least, has been a very useful tool. It's built a lot of confidence for what is required next year and has given us an opportunity to practise with the EMS in as close to a real situation as is probably possible. It has provided us with enough information for us to start planning our requirements for next year whilst also highlighting the significant changes and challenges of the new system. I've enjoyed it anyway!’*

Chapter 4
Results from national matching

This chapter presents the findings on the match rates for all LAs and VJBs, following matching with the DWP CIS system, variations across different areas and the possible reasons for that, administrator views taken from the survey on their match rates and confidence in using the confirmation process, and variations by elector type.

Green match rates for all electors

The national average green match rate for all electors was 78%. This ranged from 47% in Kensington & Chelsea to 86% in Mansfield. In addition, the median green match rate was 81%, revealing that most LAs achieved match rates towards the higher end of the range. Compared to the findings of the 2012 confirmation pilots, which found that around 70% of electors could be confirmed, this is a far better result than was anticipated.

We know from the pilots we conducted into confirmation in 2012 that some groups are less likely to confirm – students, people in privately rented accommodation, people living in communal establishment and recent home movers (there are some overlaps between these groups). In addition we know that some address types are more difficult to match due to their more complicated formatting e.g. student halls of residence. The EC have provided ward level data to academics at Plymouth University to compare with Census data on demographics and look at drivers for registration using regression analysis. The results of this can be found in the EC's Evaluation of CDR available at www.electoralcommission.org.uk.[3]

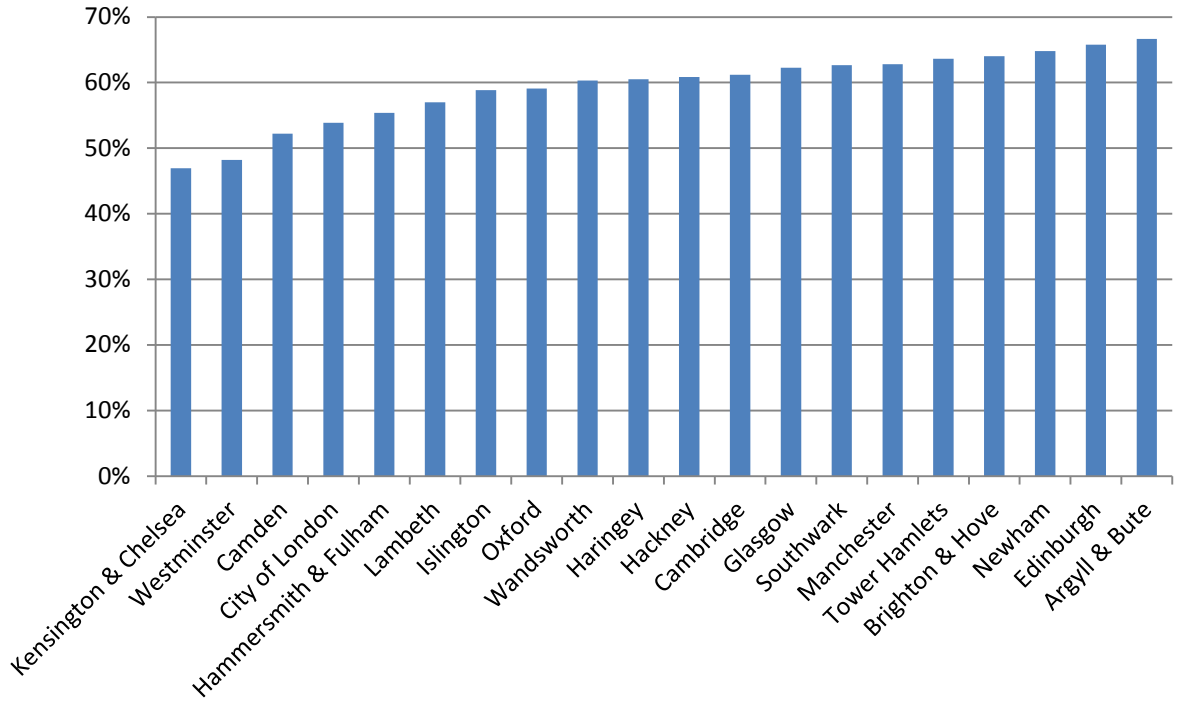
The graph in figure 3 shows the 20 authorities that achieved the lowest green match rate. Of this selection, 19 of these were predominantly urban areas, including 13 London boroughs.[4] This aligns with the previous pilot findings and suggests that it is

[3] Available from November, 2013.

[4] Local Authority classifications for England are taken from: Rural/Urban Local Authority (LA) Classification (England), Office for National Statistics. Local Authority classifications for Scotland are estimated based on information taken from: Urban Rural Classification 2009-2010 Population Tables, General Register Office for Scotland. It has not been possible to accurately identify the classifications of local authorities in Wales. An assumption has therefore been made on the basis of the LA's official

more difficult to confirm electors in urban areas due to population churn, address types and higher proportions of some demographic groups known to be less likely to confirm, for example students. The only exception within this selection was Argyll & Bute, which is predominantly a rural area. However, their low match rate can be explained by the relatively high number of second and holiday homes within the authority.[5]

Figure 3: Local Authorities with the lowest green match rate for all electors



Looking at the data for CDR, there are an additional number of reasons to explain why these authorities achieved a lower match rate. Aside from being largely urban areas, all 17 of the English authorities in this selection have a very high proportion of privately rented homes - all above the English average of 17%.[6] Equally, the proportion of people who have been residing in the UK for less than 2 years is directly linked to the success of the green match rate. All 17 of these same authorities are above the national average of 2% in this respect, indicating a high level of movement between properties in these areas.[7] These figures both indicate that addresses may have been more difficult to match, providing a reason as to why these authorities achieved a lower match rate.

title. For example, a City Council would be assumed to be predominantly urban, whilst a County Council would be assumed to be predominantly rural.

[5] According to the 2001 census, the number of second and holiday homes in the Argyll & Bute was 5,185 (http://www.scotland.gov.uk/Resource/Doc/1125/0086619.pdf).

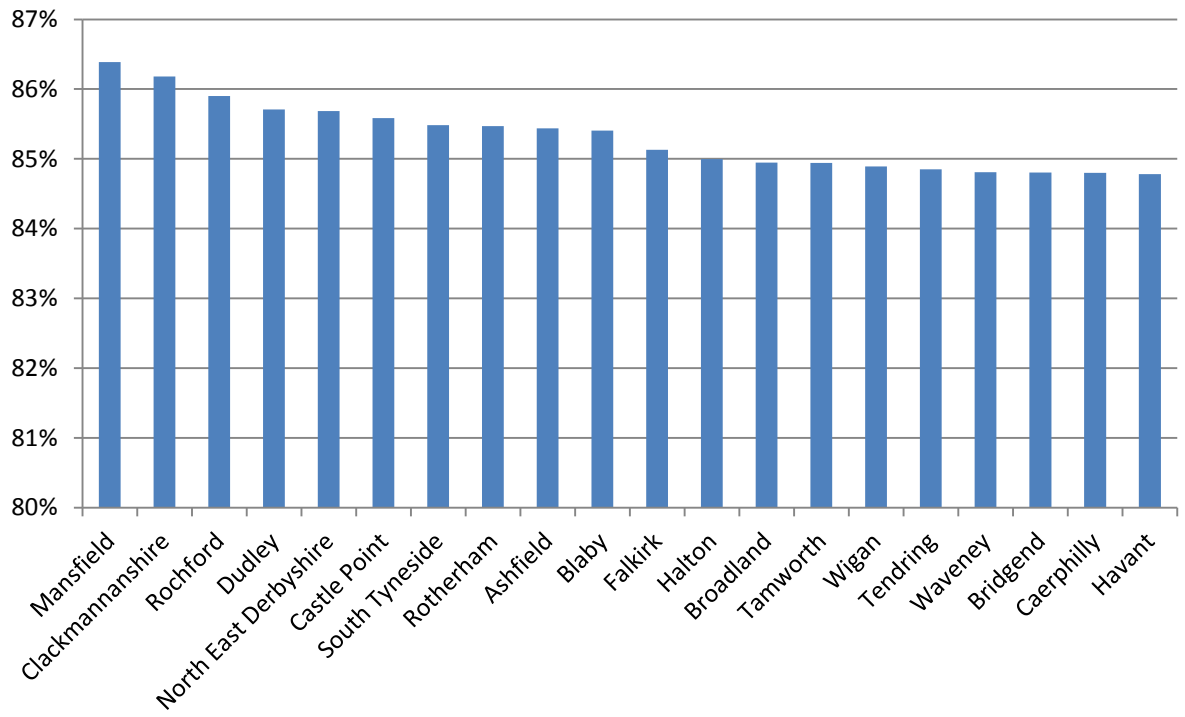
[6] 2011 Census: Table QS403EW Tenure – People, local authorities in England and Wales, Office for National Statistics (http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-and-quick-statistics-for-wards-and-output-areas-in-england-and-wales/rft-qs403.xls)

[7] 2011 Census: Table QS803EW Length of Residence in the UK, local authorities in England and Wales, Office for National Statistics (http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-authorities-in-england-and-wales/rft-table-qs803ew.xls)

In addition, these same 17 authorities all have a student population above the English average of 5% of the electorate.[8] Oxford and Cambridge in particular have a student population that is far above this average – in excess of 23% of their electorate. This shows why these areas were not as successful in confirming their electorate, as students are less likely to be registered at their university address. Moreover, most of these same authorities also have a population of 18-24 year olds above the English average of 12%.[9] In particular, the 4 English authorities outside of London (Brighton & Hove, Cambridge, Manchester and Oxford) have a population of this demographic that is over 20% of their electorate. This provides another explanation for a lower match rate as previous research has shown that people within this demographic are less likely to be registered than any other.

Similarly, figure 4 below shows the 20 authorities that achieved the highest green match rate. While authorities with lower green match rates were predominantly urban, 14 of the authorities in this selection are classified as urban while the remaining 6 are largely rural, revealing a much wider spread of LA type.

Figure 4: Local Authorities with the highest green match rate for all electors



It has already been discussed that a correlation exists between low green match rates and high proportions of privately rented housing. A similar correlation appears to exist for the high green match rates – all 16 of the English authorities in this selection fall below the English average of 17%, where authorities with a lower green

[8] 2011 Census: Table KS501EW Qualifications and students, local authorities in England and Wales, Office for National Statistics (http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-authorities-in-england-and-wales/rft-table-ks501ew.xls)

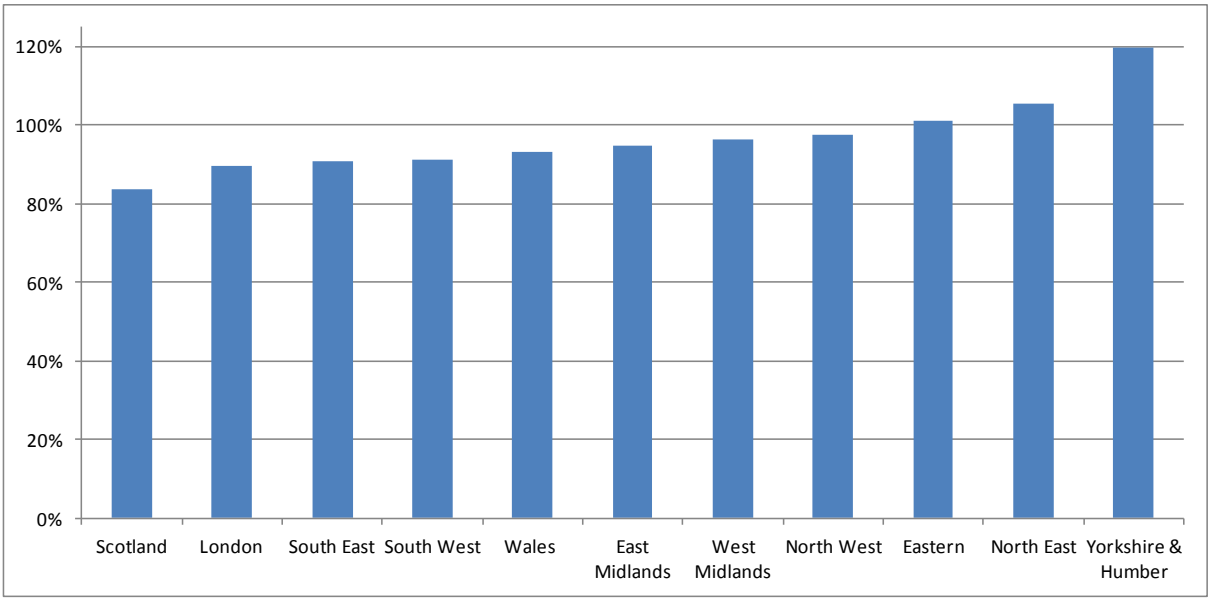
[9] Mid-2012 Population Estimates: Single year of age and sex for local authorities in the United Kingdom; estimated resident population, Office for National Statistics (http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-319259)

match rate tended to be above this average. Likewise, this analysis can be applied to the proportions of people residing in the UK for less than 2 years in these areas. The same 16 authorities are below the English average of 2%. Bridgend and Caerphilly, the 2 Welsh authorities in this selection, also follow this pattern. Both authorities fall below the Welsh average of both privately rented housing (14%) and residency in the UK for less than 2 years (1%). This shows that while authorities with lower green match rates have a higher level of movement between properties, for authorities with higher green match rates this level of movement is far lower.

Similarly, the same patterns emerge when examining the populations of students and 18-24 year olds in these areas. Again, all 16 of the English authorities in this selection fall below the English average populations of both students (5%) and the 18-24 year olds (12%). However, it should be noted that the distance from this average does fluctuate depending on the type of local authority. Dudley, for example, a predominantly urban area, has a student population of 3% and an 18-24 year old population of 10%, whilst North East Derbyshire, a largely rural area, has a student population of 2% and an 18-24 year old population of 9%. Likewise, Bridgend and Caerphilly are also below both the Welsh average student population of 5% and 18-24 year old population of 12%. The low number of electors in these demographics serves to explain why the matches against address in these areas were so much stronger, compared to LAs with a far greater number of these electors.

The lowest average unique property reference number (UPRN) match rate was in Scotland, where it was 84% (see figure 5). The second lowest was London, where it was 90%. The national average was 95%. This suggests UPRN coverage on either CIS or the electoral register, or both is lower in Scotland and London.

Figure 5: Average UPRN match rate by country / region



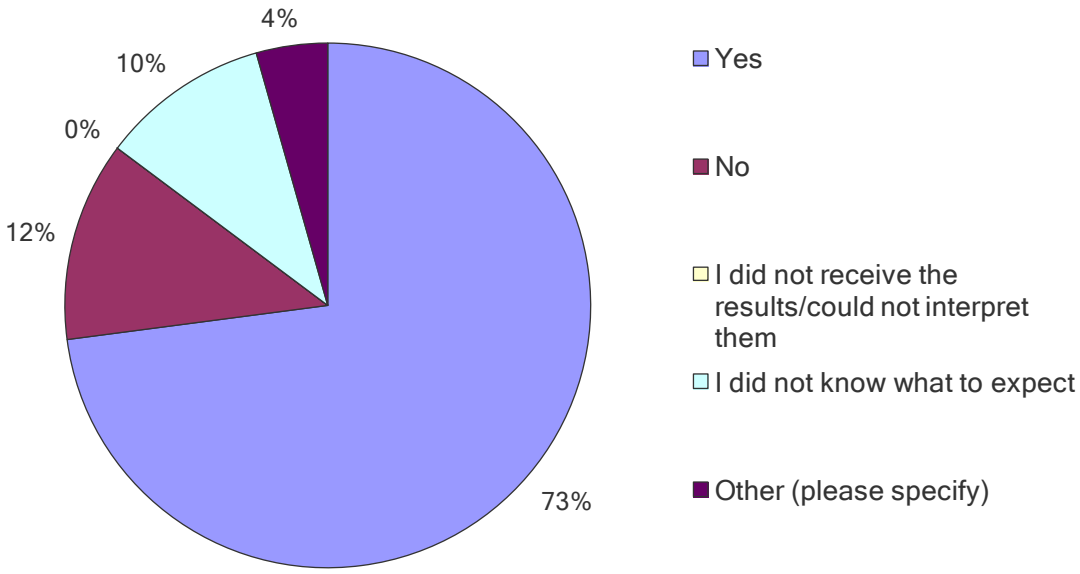
The overall average of amber address matches was 3%. 6% of address entries in Scotland were returned as amber, and London and Yorkshire & Humber both saw 4% of their addresses returned as amber. London and Yorkshire & Humber had the highest level of red address matches at 4%, with both Scotland and the South West receiving 3% of their address entries back as red. Scotland had the highest

percentage of weak address matches at 3%, with the South West at 2%. The average was 1%. All of these figures suggest address matching is more difficult in Scotland, London, Yorkshire & Humber and the South West, and least successful in Scotland and London. This supports the findings of our previous confirmation pilots, which found that addresses such as Scottish tenement buildings and sub-divided houses, houses of multiple occupation, communal establishments and flats are more difficult to match and that these areas potentially have higher proportions of those types of properties.

Views on match rates

73% of respondents felt that their overall match rate was in line with their expectations, and 10% felt that they did not know what to expect from their match rate (see figure 6).

Figure 6: Is the match rate for your area in line with your expectations?



63% of respondents felt that their match rate was better than they expected, 31% felt that it was worse than they expected, and 6% were surprised by the number of red and amber ratings that were reported among those that they considered to be certain greens. While it is pleasing that for many the match rate was higher than was expected, Cabinet Office could have more clearly communicated what the expected match rate would be to all electoral administrators, for example by trying to disseminate findings from our previous pilots more widely.

Elector types

As well as the match figures for all electors, data was also provided by LAs on the match rates broken down by elector type. These types were attainers, postal voters, proxy voters and carry-forward electors. This breakdown is useful as it reveals if and

how the match rates vary across each type and provides LAs with more specific information to help improve match rates by targeting specific groups.

Attainers

The average green match rate for attainers was 85%, 7% higher than the average match rate for all electors. This ranged from 60% in Kensington & Chelsea to 95% in Rochford. This range is both smaller in scale and higher in value than that for all electors, indicating that attainers are more likely to confirm once on the register (this echoes findings from our previous pilots on confirmation). This is perhaps largely due to the fact that they would have been more recently registered. In addition, the median match rate was 86% showing that most authorities' match rate for attainers fell towards the top half of the range.

Postal Voters

The average green match rate for postal voters was 85%. Like the average rate for attainers, this is 7% higher than the average for all electors. However, its range is from 52% in Kensington & Chelsea to 92% in Redcar & Cleveland. This range is both greater in scale and higher in value than that for all electors. The median match rate was 86%, indicating again that most electors in this category fell towards the higher end of the range.

Proxy Voters

The average green match rate for proxy voters was 77%, 1% lower than the average for all electors. This ranged from 31% in North Devon to 100% in Ashford, Babergh, Brent, Brentwood, Chelmsford, Corby, Epping Forest, Epsom & Ewell, Fylde, Greenwich, Newport, Oadby and Wigston and Warrington. This is a far greater range than the other elector types. However, like the other elector types, the median match rate of 78% indicates most authorities fell towards the top end of the range.

Carry-forward electors

The final elector type included in data provided by authorities were electors 'carried forward' from the previous register. This group differs from the other elector types, as when confirmed they will not be transferred automatically onto the IER register unless they have been included on a Household Enquiry Form (HEF) as part of the IER canvass. It was discussed in the evaluation of the 2012 confirmation pilots that this may impact on the overall confirmation rates. It is therefore important to review the match rates for this group of electors to ascertain how they impact on the overall confirmation rates.[10]

The average green match rate for carry-forward electors was 53%, ranging from 5% in Bolsover to 85% in Clackmannanshire. With a median match rate of 55%, this shows that most authorities fell towards the lower end of the top half of this range. This is clearly a much lower result than that for all electors and the other elector

[10] Not all files included accurate figures for carry-forward electors as data was not available for all local authorities due to problems some had in their automated reporting.

types. We would expect this as by their nature, carry-forward electors details are less likely to be current given they have not been checked in longer.

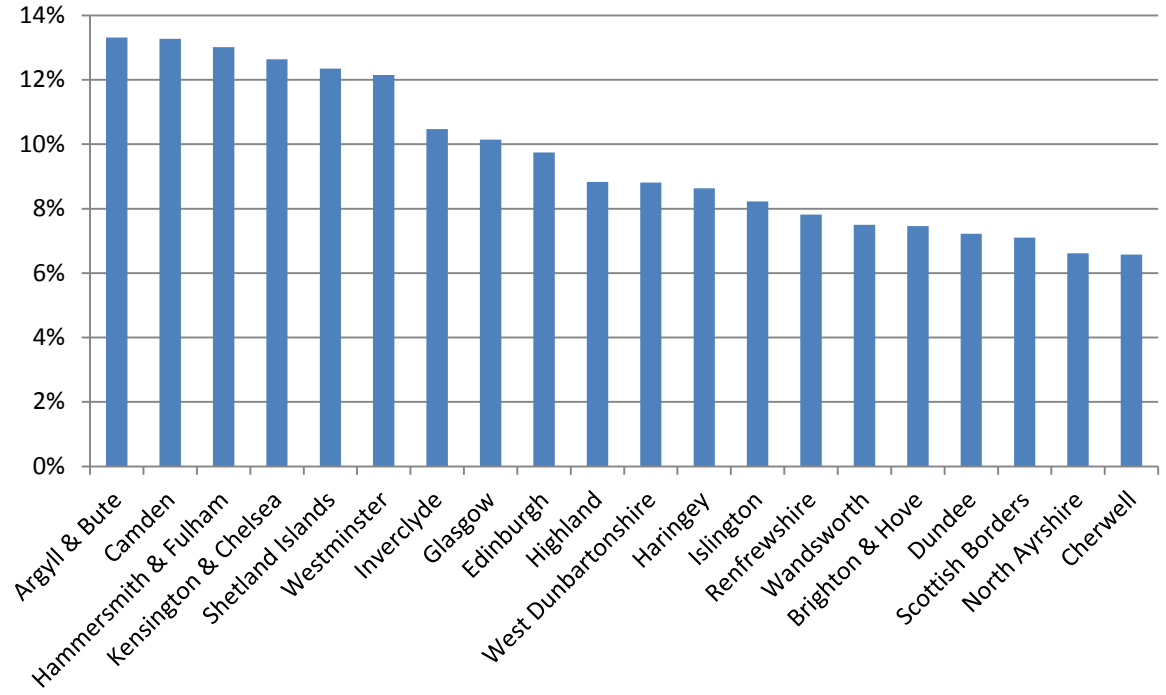
When carry-forward electors are removed from the data set, the overall average green match rate is 76%. The difference that carry-forward electors make to the green match rate ranges from 0% in Bolsover to 11% in Bolton. On average this makes 2% difference to the green match rate. Whilst not significantly lower, this does show how this elector type's exclusion from the overall data set can impact on the overall confirmation rate. In live running we would expect some of these electors to be captured by HEFs and therefore be treated as confirmed so the impact on the overall match rate will be smaller. The range of the overall average green match rate excluding carry-forward electors is between 46% in Kensington & Chelsea to 86% in Rochford. This is a largely similar result to the overall match rate range when this elector group is included. In addition, the median match rate of 79% is 2% lower than when carry-forward figures are included.

Amber match rates

In addition to the green match rates that have already been widely discussed, the amber match rates have also been analysed. These figures are useful when compared to lower green match rates as they can indicate how close certain local authorities came to confirming more electors, and they may be easier to confirm using local data matching than red records. The national average amber match rate for all electors was 3%, which ranged from 1% in Tamworth to 13% in Argyll & Bute. The median amber match rate was 3%, indicating that most amber matches fell towards the lower end of the range.

The graph in figure 7 shows the 20 authorities that achieved the highest amber match rate. Several of the authorities that achieved lower green match results have appeared in this selection. All 7 of the London boroughs included, Argyll & Bute, Brighton & Hove, Edinburgh and Glasgow demonstrate this point. Of these 20, 15 are classified as urban with the remaining 5 as rural. It is also interesting to note the high number of Scottish authorities in this selection – over half with 11 in total, 7 of which are predominantly urban and the remaining 4 of which are predominantly rural. As was discussed earlier, Argyll & Bute achieved a low green match rate of 67%. It was argued that this was due to the high number of second and holiday homes in the area, and this suggestion can also explain the relatively high amber match rate. We know from our previous pilots that certain types of addresses are more difficult to match – communal establishment, sub-divided flats and Scottish tenement buildings, and so these areas with higher amber matches are likely to have higher proportions of those property types.

Figure 7: Local authorities with the highest amber match rate for all electors



Actions based on match status

What would you intend to do with the entries marked as green?

Confirm them all on the register	81%
Confirm some on the register	9%
Confirm none	0%[11]
Don't know	3%
Other	7%

81% of respondents will confirm all of their green matches on their registers. Of those who will confirm some of the entries, 55% will carry out local data matching and send out invitations to register as necessary, 14% will send out an invitation to register without carrying out local data matching, and 22% answered that they would perform a number of other activities, such as using local data matching if they have funds to purchase data matching software. 7% of respondents answered that they would do something else with their entries marked as green: in most cases, the respondent needed more information before they can state what they will do with green entries; in others, the respondent intends to do local data matching with specialist software, but only if their budget permits.

What would you intend to do with the entries marked as amber?

[11] A number of authorities answered that they would not confirm any green matches. Discussion with these respondents has shown that they understood the question to be asking them whether they would check if green matches were indeed green, to which they answered that they would not. They will in fact send confirmation letters to all green matches.

Carry out local data matching with them	86%
Send them an invitation to register without carrying out local data matching	3%
Don't know	2%
Other	9%

ESMs are required to take further steps to verify the entitlement to remain registered of those entries which are marked as amber.[12] 86% of respondents will carry out local data matching on their amber records, and 3% will invite them to register without carrying out local data matching. Of the 9% who indicated that they would carry out a different action to those listed, a number stated that they would carry out local data matching on some of the records and send out invitations to register to others; others stated that they do not know what they will do at this time; and others stated that they will carry out local data matching if their budget allows them to purchase specialist software to do so. Local data matching should increase the overall match rate in that authority.

What would you intend to do with the entries marked as red?

Carry out local data matching with them	74%
Send them an invitation to register without carrying out local data matching	11%
Don't know	4%
Other	12%

ESMs are required to take further steps to verify the entitlement to remain registered of those entries which are marked as red.[13] 74% of respondents indicated that they would carry out local data matching on those entries marked as red, while 11% would invite them to register without carrying out local data matching. Of the 12% who indicated that they would carry out a different action to those listed, a number stated that they would carry out local data matching on some of the records and send out invitations to register to others; others stated that they do not know what they will do at this time; and others stated that they will carry out local data matching if their budget allows them to purchase specialist software to do so. Local data matching will increase the overall match rate in that authority.

Confidence in using confirmation

[12] As required by article 7(7-9) of The Electoral Registration and Administration Act 2013 (Transitional Provisions) Order 2013

[13] As required by article 7(7-9) of The Electoral Registration and Administration Act 2013 (Transitional Provisions) Order 2013

86% of respondents are very or fairly confident about using the confirmation process tested in the CDR to passport electors onto the new register during the transition to IER. However, 10% of respondents did not feel confident. The largest concern is with the quality of DWP's CIS data, with respondents stating that they believe their own local data is more accurate. Respondents are consequently worried about the number of red matches on electors that they believe should be green – a number gave examples, for instance, of individuals who have been resident at the same property for a number of years and who have been returned as red.

The Cabinet Office confirmation pilots in 2012 specifically looked at the accuracy of data matching for the purposes of confirmation. In order to do that each participating area was asked to provide two versions of their electoral register to be matched against DWPs CIS – their pre-canvass register (just before the annual canvass) and their post canvass register (just after the annual canvass), the results of which were then compared against each other. The reason for undertaking this exercise is because the completeness and accuracy of the register declines during the electoral cycle. Therefore by comparing the results for the two registers (when they are likely to be at their least and most complete) it is possible to assess the proportion of individuals positively matched against DWP who were subsequently confirmed as resident at the same address during the annual canvass. This gives an indication of the accuracy of the matching including the potential level of inaccuracies that might be caused by population churn. Analysis of the pilot results showed that the risk of inaccuracies were small[14]. DWP data is reliant on individuals providing updates of any changes in their circumstances such as moving home or changing their name, or having a recent interaction with DWP which then leads to their information being updated on CIS. This means that whilst the confirmation process can confirm the majority of electors, there will be some who cannot be matched, even when their information may be accurate on the electoral register.

We are confident in the accuracy of the matching process, but need to ensure that electoral administrators feel confident in the data and have realistic expectations about the matching process. This is a communication issue which we intend to address via a combination of Cabinet Office communications as well as specific guidance on the confirmation process. In addition, for those who are not confirmed local data matching can increase the match rate and others will be captured through the 'write out' process administrators conduct post-confirmation.

A number of respondents also reported cases of receiving reports in which electors were marked as 'deceased' although they believed they were still alive. Examples of this issue have been sent by the Cabinet Office to DWP for further investigation. The problem may lie with one element of the matching algorithm – Soundex – which might cause electors to be incorrectly flagged as deceased because it sounds similar to the name of a deceased person (and not because the person is actually labelled as deceased on DWP's data). Cabinet Office are currently looking into how the match reports can show whether a deceased flag has been issued based on a unique match, or a Soundex match alone. Some respondents are concerned about the cost of writing out to unmatched individuals, both in terms of staff time and write-out materials. Some feel that the match process is overly complicated, that mixed

[14] See pg 17 of Cabinet Office Evaluation and Annex D available at: https://www.gov.uk/government/publications/simplifying-the-transition-to-individual-electoral-registration

RAG households are confusing, and that the entire confirmation process will generate confusion among electors. The EC's Guidance, which includes Cabinet Office Ministerial Guidance, aims to provide clear guidance on the confirmation process for service managers to tackle the first two issues, and the EC's large-scale national publicity campaign will mitigate the latter. It is also important to remember that confirmation has been implemented to free up the resources of administrators so that they only have to focus on registering the minority of electors who have not been confirmed.

Wards

In addition to the data described above, LAs were also required to provide details of their RAG matches broken down by individual wards. The data has been provided by the EC to academics at Plymouth Universtiy who had used it and the Census data to conduct regression analysis to identify key drivers for confirmation i.e. which demographic groups are more or less likely to confirm. This analysis can be found in the EC's Evaluation of CDR available at www.electoralcommission.org.uk.[15] This data shows how these matches varied across different areas within the authorities, and highlights where particular wards may have brought down the overall match rate for an authority. This is useful as it helps authorities target specific areas and groups when looking to improve their overall match rates.

This point is demonstrated by the match results achieved by York City Council. York achieved an overall average green match rate of 74%. It's range, however, was between 11% in Heslington and 87% in Haxby and Wigginton. It is almost certain that Heslington's low match rate is due to the fact that it is the site of the University of York and therefore an area with a highly concentrated student population. This reiterates the assertion that there is a direct relationship between match rate and student population. Equally, Oxford's ward data shows a similar pattern. The average match rate was 59% which, as has already been discussed, is largely due to the high student population residing in the city. However, this ranges form 79% in Northfield Brook to 7% in Holywell. An explanation for the exceedingly low match rate in Holywell might be it's relatively high population of 18-24 year olds.[16]

The ward data also shows how match rates can vary according to the type of geographical environment. Camden's average match rate was 52%, ranging from 36% in Bloomsbury to 69% in Highgate. Bloomsbury is in the very centre of London and is the site of University College London, and is therefore likely to have both a highly concentrated student population and a high percentage of movement between properties. By contrast, Highgate is a more suburban area of the authority, showing how a geographical environment such as this can affect the match rate. Manchester demonstrates a similar variation – it achieved an average match rate 63%, whilst the green match rate in the City Centre ward was just 25%.

[15] Available from November, 2013.
[16] http://openlylocal.com/wards/7885-Holywell.

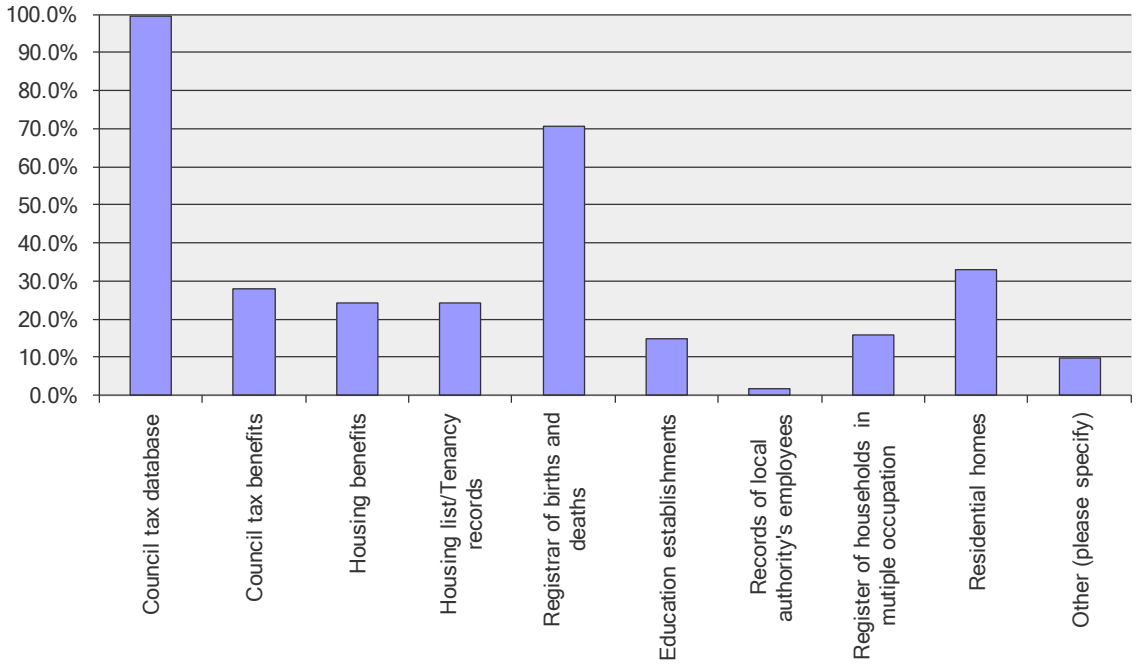
Chapter 5
Local data matching results

During the transition to IER, EROs will be able to use local data matching to confirm individuals who could not be matched against the DWP database, adding to the overall match rate in their area. Whilst this was not a compulsory part of the CDR, LAs were invited to use locally held data sets (for example Council Tax data or Housing Benefit data) where they had capacity to do so to see if they could improve their overall match rate.

As shown in figure 8, 97% of respondents to the survey normally use local data as part of their annual canvass activities. Of those 97%, nearly all use council tax records, 70% access the registry of births and deaths, 30% view records held by residential homes, and others use benefits records, housing lists, records from educational institutions, HMO registries, the 'Tell Us Once' system[17], the Local Land and Property Gazetteer[18], health and safety lists, and recycling records.

[17] A system used by many authorities in which a resident completes one form to inform the authority of a birth or death and the details are entered into all the relevant local databases.
[18] The address database maintained by authorities.

Figure 8: Which local databases do you normally use as part of your canvass activities?

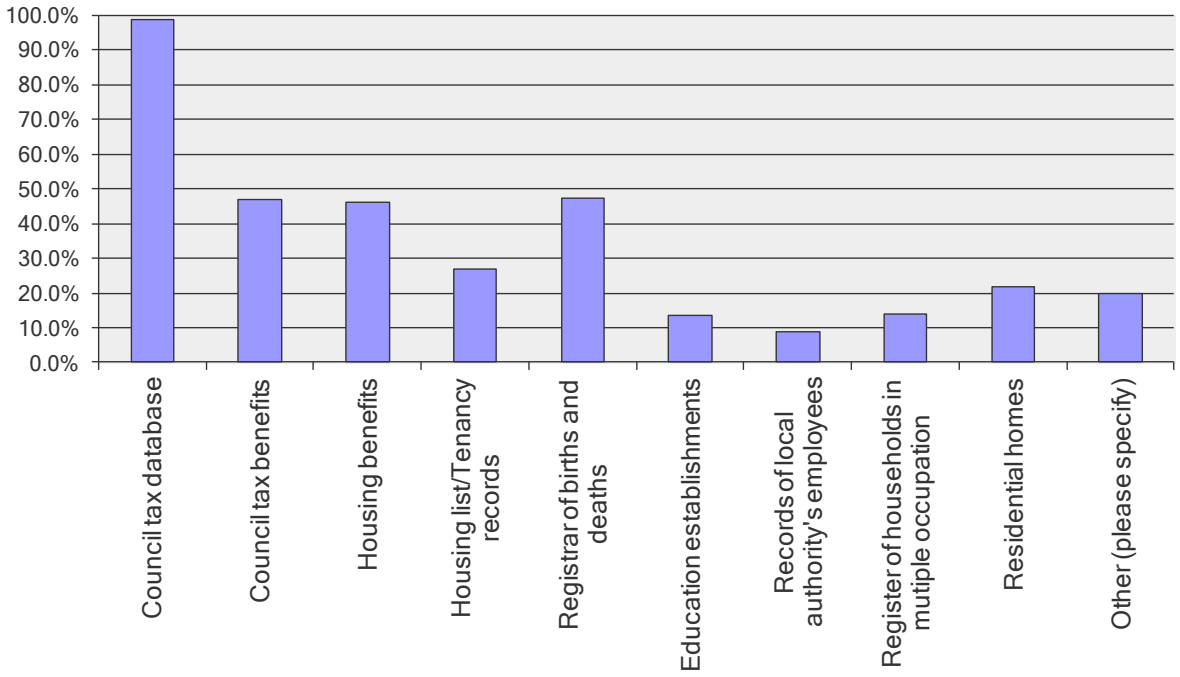


87% of respondents conducted local data matching on the records returned to them by the dry run. Of those that did not conduct data matching, 67% did not have the staff resources to carry out data matching at that time and 23% felt that their match rates were high enough that they did not need to carry out local data matching. Other responses included the view that authorities did not have the resources to conduct data matching during a busy holiday period or that the authority will automate data matching in the future by purchasing a software enhancement to do so. Some authorities did not see the need to carry out data matching during the CDR when the register will naturally change over the course of the coming year before confirmation itself (though it should be noted that conducting local data matching at this time could be seen as useful in giving an indication of potential workload in live running).

68% of respondents followed the same processes with regards their local data during CDR as they do during normal annual canvasses; 33% followed new and different processes. Of that latter group, 18 respondents are developing new, automated data matching processes, 17 have developed new, manual data matching processes, 12 have already purchased and are using data matching software, 11 compared their downloads against more records than normal and 4 had no time to analyse their CDR downloads.

As figure 9 below shows, nearly 100% of respondents who carried out local data matching during the CDR used the council tax database. 47% made reference to the register of births and deaths, 46% used benefits lists, 26% accessed housing lists, and 22% looked at registers of residential homes. Other sources consulted include the Local Land and Property Gazetteer, the 'Tell Us Once' service, the National Fraud Initiative database, leisure and library service records, parking permits service, the adult social care lists and bus pass records.

Figure 9: Which local databases have you / do you plan to use for matching as part of the dry run?



24% of respondents intend to complete local data matching manually by 'eyeballing' records. 22% will do it automatically, and the remaining 54% of respondents will carry it out by using a mixture of manual and automatic checking.

Whilst we do not know how many electoral administrators will chose to conduct local data matching during confirmation in 2014, 91% of respondents believe that local data matching will be a very or fairly important part of the live confirmation process. This means that the overall match rate during live running is likely to be substantially increased.

138 of 380 local authorities conducted local data matching on their CDR results *and* submitted data in an automated report from their EMS system to the Cabinet Office. Results from this data suggest that local matching adds an average of 7% to the overall green match rates. This ranged from 0% in Wyre and 34% in Hammersmith & Fulham, with a median of 7%.

The results show that local data matching was particularly helpful in Scotland, southern regions of England, London and in authorities that achieved lower green match rates. Hammersmith & Fulham is a good example of this. They achieved a green match rate of 55%. However, local data matching was able to add 34% to this match rate. This changes their green match rate to 89%, placing them in the higher region of the overall green match rate.

Chapter 6
Workloads for administrators in live confirmation

Confirmation

There is concern among administrators that confirmation creates a large workload. Their worries cluster around the following areas:

1. Co-ordinating with local IT services to ensure that the right IT systems are in place to allow confirmation to go ahead
2. Reading and understanding documents sent to them by EMS suppliers, the Cabinet Office and the EC
3. Allocating staff time to monitor the up- and download
4. Receiving and interpreting the results
5. Carrying out local data matching
6. Sending out confirmation letters, HEFs, and Invitations to Register (ITRs)

If respondents are not carrying out local data matching, it is most commonly because they feel that time and budget constraints make it impossible. The majority of respondents who commented feel that confirmation itself is an additional burden created by the transition to IER. It should be made much clearer through Cabinet Office and EC communications that although confirmation creates new challenges – coordinating IT services, for example – it *reduces* the amount of canvassing work that needs to be done and passports an average of 78% of electors onto the new IER electoral register.

LAs were asked to provide details of the letters that would be sent on the basis of their match results, both before and after local data matching was carried out. This is helpful as it enables us to see the impacts local data matching will have on ERO workload during confirmation live run in 2014. The results below discuss this in further detail.

The 3 types of letters that would be sent are:

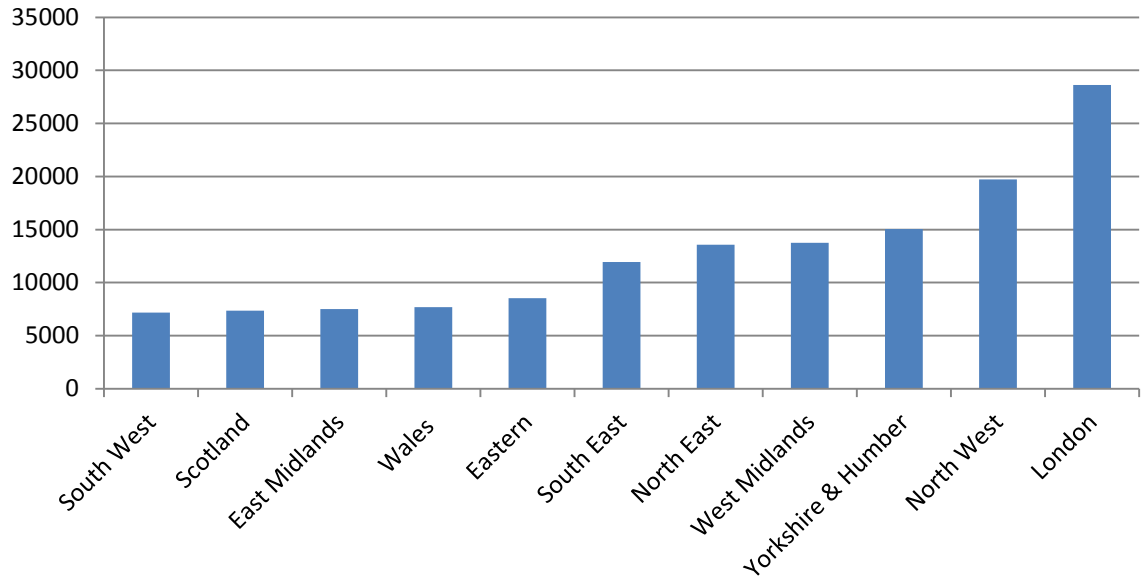
- Confirmation letters, sent to electors who have been confirmed
- ITRs, sent to electors who have not been confirmed inviting them to register
- HEFs, sent to households where there are no electors registered and where the electoral adminstrator believes there may be new or additional electors to try and gather their information and add them to the register. (The reporting tool generally counted HEFs that were sent to households where there had been individuals on the register and not empty/'void' properties as these records had not been sent for matching. However, HEF figures including void properties were requested in the feedback survey)

Overall, local data matching reduces the number of ITRs and HEFs that would need to be sent by an average of 21%.[19] This ranged from a 96% decrease in the number of letters in Halton to a 0% change. The median was a reduction of 19%. This indicates that local data matching can successfully reduce the workload on EROs when issuing letters.

HEFs

Based on data provided by 255 authorities, the average number of HEFs which will need to be sent (including to empty or void properties) is 12,874 per authority. As shown in figure Figure 1010, the Greater London authorities reported needing to send the highest number of HEFs on average, at 28,619. This is likely to be due to the higher level of churn in properties in London. The South West and Scotland reported needing to send the lowest number of HEFs on average, at 7,182 and 7,353 respectively. The median number of HEFs to be sent is 7,487.

Figure 10: Average number of HEFs to be sent by each office in region / country



[19] These figures are based on 102 local authorities who provided information on the numbers of letters that would be sent both before and after local data matching.

Chapter 7
Summary and conclusion

Cabinet Office pilots originally conducted in 2011 identified the possibility of data matching the electoral register against DWP CIS data to confirm electors to a level at which we could feel confident that their name and address remained the same. They could therefore be 'passported' across to the new system without electoral administrators having to re-register them, freeing up resources to focus on the minority of electors who have not confirmed and encouraging those missing from the register to join. In 2012 further pilots were conducted to specifically test data matching for confirmation and to check the accuracy of the data and matching process. The pilots took place in 14 areas and saw a match rate of around 70% (an improvement on the 66% seen in the original pilots). The pilots noted significant variation across different local areas which were driven by population characteristics and issues with addresses. The pilots found that we could be confident in the accuracy of this matching.

The CDR exercise that took place between July and October 2013 was an opportunity to test confirmation across all LAs and VJBs in Great Britain with a fully enabled IT system. This therefore allowed us to assess how confirmation might work under 'real' conditions when it is used in the summer of 2014 and allowed electoral administrators to have a better understanding of their possible match rate and their related workload and necessary resources for 2014. The match rates form part of the funding formula that the Cabinet Office will use to award funding to LAs and VJBs for any additional burden that IER will place upon them.

Key findings and implications

Overall, the results of the CDR are better than expected with an average match rate of 78% and a median rate of 81%. Local data matching has shown the potential to add on average a further 7% to the overall match rate. Feedback on the process of CDR was largely positive, but some lessons and possible improvements have been identified for the Cabinet Office to work on ahead of confirmation in 2014.

Views on the process of conducting the CDR

Most administrators who responded to our survey were satisfied with the guidance they received in preparation for CDR but a reasonable number would like one source of guidance and some asked for more detailed guidance on how to conduct local data matching. Both of these suggestions have been addressed by Ministerial Guidance which was incorporated with the EC's own guidance and published in October 2013. In terms of the IT, the vast majority of survey respondents found it easy to upload their registers for matching and were satisfied with the way their match report was downloaded back into their EMS system. However, there were some problems with some duplicate records being uploaded and some found that it took a long time for their results to be downloaded. There was also a presentational issue with the matching algorithm in the early stages of CDR with records where the elector could not be matched automatically changing the address RAG to red; this was resolved by the Cabinet Office, the affected local authorities were all given the opportunity to re-run their register, and their overall match rates were not affected.

Most administrators found it easy to analyse their data but there is a varied skilled set for this type of work amongst the electoral community and some relied on support from their IT department. Some were also unsure of the reasoning behind the matching and why some individuals might fail to be matched. From a matching perspective this is difficult to resolve – there could be a large number of reasons why an individual record has a red or amber RAG – so instead there is a need to improve understanding of the matching process and manage expectations of administrators. Some also queried the accuracy of DWP data. While we accept that some records will be inaccurate and/or out of date, testing from the pilots in 2012 showed that we can be confident in the accuracy of the matching and so again there is a need for the Cabinet Office to reassure administrators about the quality of both the data and the matching process to build their confidence.

However, overall, most administrators said they would be confident in using confirmation results, would treat green matches as confirmed and issue them confirmation letters, carry out local data matching on amber and red records to see if they could increase their match rate, and then write out to those they were unable to confirm.

DWP match results

The national average match rate was 78%; this ranged from 47-86%, with a median rate of 81% – showing that most local authorities had a match rate towards the higher end of the spectrum. This is a better rate than anticipated prior to CDR and shows that the majority of existing electors will be confirmed in 2014. We know from our previous pilots that some groups are less likely to confirm – students, people living in privately rented accommodation, people living in communal establishments and recent home movers (there are clearly some overlaps between these groups). In addition, we know that some address types are more difficult to match due to their more complicated formatting e.g. rooms in student halls of residence. These findings were replicated in the CDR with 19 of the 20 areas with the lowest match rates being predominantly urban areas, including London boroughs where there is a high churn, lots of flats and sub-divided properties and a high proportion of privately rented flats, and areas with high proportions of students such as Oxford and Cambridge.

As with our previous pilots, some groups seemed more likely to confirm once they were on the register – notably attainers. Academics at Plymouth University have conducted regression analysis using the ward level data on matching and data from the Census to look at which demographic groups are more or less likely to confirm. Results can be found in the EC evaluation of CDR, available at www.electoralcommision.org.uk.[20]

If an electoral administrator has reason to believe that an elector is still at a property but hasn't responded to the canvass, they can choose to 'carry-forward' that elector and keep them on the register. During live running of IER in 2014, electors who are confirmed but who are carry-forwards will not be treated as confirmed in the same way as other electors (because of the possibility that their details may be less current and accurate). Instead their residence will be sent a HEF and if they are named on that form they can then be treated as confirmed. It is not possible to know how many of the confirmed carry-forward electors will be truly confirmed in live running by replying to a HEF but we were interested to assess what the match rates for carry-forward electors were during CDR and the effect of this on the overall match rates.

The average match rate for carry forward electors was 53%, ranging from 5-85% and with a median of 55%. This is much lower than for non-carry forward electors, which is to be expected given that their details on the register are more likely to be out of date or inaccurate due to their lack of response to the annual canvass. When carry-forward electors are removed from the dataset the overall match rate falls to 76% but we would expect many of these to confirm via a HEF and therefore do not expect to see a large effect on the overall national match rate during live running in 2014.

Local data matching

Most respondents to our survey said that they conducted local data matching as part of their normal canvass activities and also did it for CDR. The most commonly used source of data for both was council tax information. Based on the reports submitted to the Cabinet Office (138 local authorities), local data matching has the potential to add an average of 7% to the national match rate. This ranged from 0-34%, with a median rate of 7%. The use of local data matching seemed particularly helpful in areas with lower DWP match rates, as might be expected.

Recommendations

Based on these findings, the Cabinet Office will look into the possibility of providing a breakdown of confirmation match results by polling district, a separate list of mixed RAG households, and percentages in the download reports wherever they are meaningful. The Cabinet Office have already provided (in the recently-published Ministerial Guidance) information in plain English on how RAG statuses are awarded, what an amber status means, and how to conduct local data matching.[21] It will

[20] Available from November, 2013.
[21] Available at http://www.electoralcommission.org.uk/__data/assets/pdf_file/0009/162576/Part-3-The-transition-to-IER-in-2014-15.pdf and http://www.electoralcommission.org.uk/__data/assets/pdf_file/0011/162578/Part-4-Maintaining-the-register-throughout-the-year.pdf.

continue to monitor feedback on this guidance and seek to provide more information to EROs where necessary.

Annex A – Data Matching methodology

The initial matching of the data was undertaken by DWP using a matching algorithm created by their Information, Governance and Security team. The database against which the electoral register is matched is the DWP Customer Information System database. CIS is an amalgamated data source, consisting of information received from internal DWP heritage systems, as well as other government sources, such as HMRC. As a result CIS is seen within DWP as being the master of customer information.

The source CIS database is updated daily, however for the purposes of the pilot matching was undertaken on snapshots of the data extracted at a similar time to the electoral registers to ensure comparability.

The DWP matching algorithm works like a filter, the stages of which can be broadly summarised as:

1. Some data standardisation of both the electoral register and the DWP is undertaken to make the records more consistent (e.g. Str. or St. to Street).
2. The algorithm searches for a matching address in the DWP data. In the first instance through Unique Property Reference Number (UPRN), then, if a match cannot be found through UPRN, by comparing the lines of the address.
3. If an address can be located the algorithm then compares the name fields to those of the individuals held in the DWP CIS database at that address.

The end result is a series of match statements that describe the levels at which a record has 'passed' or 'failed' the different matching criteria.

It is on the basis of these statements that the scoring algorithm was then applied, assigning a RAG category to the address, identity and then applying an overall RAG. This part of the process will be part of the IER DS in live running, and was carried out by GDS for the purposes of the pilot.

The RAG status was assigned according to the following criteria:

Address RAG

Any DWP address match was assigned a 'green' RAG. The minimum match criteria for an address is complete postcode + the numeric from address lines one/two. Details were also provided to the ERO as to whether the address match had been made using the UPRN or through a straight address match.

Identity RAG

A positive 'green' identity match was assigned to records that matched at any one of the following levels:

- Full first name **plus** full last name

- Full last name **plus** first three initials of first name
- As above but including middle name/initial/DOB where available and matched
- Date of birth (DOB) **plus** full lastname
- Full last name **plus** soundex[22] match on first name **plus** either middle name **or** middle initial **or** DOB
- Soundex match on last name plus first name **plus** middle name **or** middle initial **or** DOB
- Full last name and full first name but reversed

A possible 'amber' identity match was assigned to records that did not match at the levels above but did match at one of the following levels as minimum:

- Soundex match on last name **plus** DOB **or** first three initials of first name
- Full last name **plus** soundex match of first name

The experience of the pilots was that the vast majority of records where a positive identity match could be found were matched on at least the full last name and full first name (approx 95%).

Overall RAG

The final match rate was calculated using the overall RAG, which in most cases was the same as the identity match, apart from the following exceptions:

a) Where DWP had a record of the individual being deceased the overall rating defaulted to 'red'

b) Where DWP had a record of the individual being older than 100 the overall rating defaulted to 'red'[23]

c) Where the DWP match has returned more than one 'best match' for a record with a property the overall rating defaulted to 'amber.'

As detailed in the methodology section, feedback from the pilot areas was used to inform the development of both the matching and scoring algorithm, including working with five Beacon pilots at the start of the pilot to refine the algorithm and the presentation of the match files.

Refined algorithm – overview of changes

For the purposes of the pilot the matching/scoring algorithms were frozen at the point of the pre-canvass matching to ensure that the pre and post canvass registers could be accurately compared. However, based on the feedback from areas obtained during the pilots some further refinements and testing of some of the elements of the algorithms was undertaken. The key elements are summarised below:

[22] A soundex match is a match made using a phonetic algorithm for indexing names by sound, as pronounced in English, so that they can be matched despite minor differences in spelling. The SOUNDEX algorithm is seen to English biased and is less useful for languages other than English.

[23] This age was set as a proxy for the pilots, in go-live this will be aligned with information known about the oldest living person.

a) Address matching - refinements to the matching algorithm to try and widen the address matching criteria. Whilst, in the pilot areas the vast majority of records (96%) did pass the initial address matching stage[24] feedback also suggested that some specific address types e.g. sub-divided properties were less likely to match. Therefore DWP explored ways to loosen the criteria for the address matching in an attempt to enable more addresses to be matched. This was done by allowing records that could not be matched at other stages to pass the address match stage where a combination of the postcode and full last name matched the electoral register record. Testing undertaken across 6 pilot areas, which included individually reviewing 823 records matched on this criteria found that 78% of these matches were deemed accurate. Whilst this suggests that the majority of these records can be correctly matched it remains a significantly lower accuracy level than other matches and therefore, in the refined algorithm, anything matching on this address category is defaulted to an 'amber' overall RAG.[25]

b) Multiple matches – as detailed above the pilot algorithm gave any records that had been matched with more than one record on the DWP database a default rating of 'amber', however a number of pilot areas queried whether this was the correct approach as they did not feel that the fact that more than one DWP record could be matched against the register entry weakened the validity of the match. Testing undertaken across 6 pilot areas, which included individually reviewing 446 records matched on this criteria, found no significant difference in the accuracy of the matching of these records compared to other positive matches. Therefore for the refined algorithm the overall RAG for these records will no longer default to an overall RAG of amber.

c) Last name plus first three initial matches – these records were given a 'green' in the pilot algorithm, but some areas expressed uncertainty about this classification as there is the potential for inaccuracies, for example "Stephen" and "Stephanie" could be matched under this category despite being different people. Testing undertaken across 6 pilot areas, which included individually reviewing 675 records matched on this criteria found that 94.4% of these matches were deemed accurate. Whilst this was a slightly lower result than other 'green' match categories in the pilot algorithm (98.4% of which were deemed accurate based on individually reviewing 1,062 records) the results suggest that including this category (on which approximately 3% of register entries matched in the pilot) would not weaken the accuracy of the matching significantly and therefore these records continue to be assigned a 'green' RAG in the refined algorithm.

In addition, DWP have made some additional changes to the pilot matching algorithm for the 'refined algorithm' including:

[24] Range 92-98%.

[25] Two additional stages of address matching which further weakened the address match (for example by using post code stub as opposed to full postcode) were also tested but the results showed that the majority of these matches were inaccurate and therefore they were discounted.

- The standardisation aspects of both name and address fields have been extended to incorporate additional strings such as *'Apartments'*, *'Buildings'* as well as taking account of the variation in non-English names such as potentially 10 variations of Mohammed
- The address detail matching has been extended to incorporate the fact that either within the register or within the DWP CIS data there is a potential for relevant information to be present within the Address line 3 data. This is particularly relevant for places such as care homes, where the first line of the address may contain the name of the care home, which then pushes the content of the physical address further across the address fields.
- The extract of DWP CIS data matched against excludes individuals under 16 years of age who would not be eligible to register.